Physica A 674 (2025) 130770

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Cryptocurrency in global dynamics: Analyzing the Crypto Volatility Index and financial markets with machine learning

Susanna Levantesi [a], Gabriella Piscopo [b,*], Alba Roviello [b]

[a] Department of Statistics, Sapienza University of Rome, Viale Regina Elena 295, 00161 Roma, Italy
[b] Department of Economics and Statistical Sciences, University of Napoli Federico II, Via Cupa Cinthia 21, 80126 Napoli, Italy

ARTICLE INFO

*Keywords:*
Cryptocurrency
Machine learning
Non-linear dynamics
Volatility

ABSTRACT

Accurate estimation of cryptocurrency market volatility is crucial for investors. The Crypto Volatility Index (CVI) was developed to measure the market's expectations for the 30-day implied volatility of Bitcoin and Ethereum to address the growing demand for reliable predictions. This study explores the relationship between the CVI and the volatility of traditional financial markets, including the Gold Volatility Index (GVZ), the Crude Oil Volatility Index (OVX), and the S&P500 Volatility Index (VIX). Three other variables are also analyzed: the USD to EUR exchange rate (USDEUR), the Federal Reserve interest rate (FED), and the NASDAQ index. The aim of the research is explanatory: the input variables and the CVI are observed contemporaneously to catch the complex relation between them. Using Pearson correlation, distance correlation, and mutual information, we demonstrate the presence of non-linear relationships between some variables in the dataset. Explanatory analysis is conducted using machine learning techniques, specifically the Random Forest (RF) algorithm and Gradient Boosting Machines (GBM) to account for these potential non-linear interactions. These methods are better suited than standard linear models for identifying complex relationships. In particular, the RF algorithm reaches a better level of accuracy than GBM and avoids overfitting.

## 1. Introduction

In recent years, the financial sector has experienced a strong remodeling due to the increasingly broader onset of digital currencies. Despite their recent design, cryptocurrencies are starting to be considered a significant asset for portfolios and trading processes. Given their growing influence and significant impact on global markets, a substantial scientific effort is being sustained to analyze the consequences of cryptocurrency appreciation from multiple perspectives, employing a wide array of techniques.

Recently, great attention has been paid to the part cryptocurrencies can play in portfolio diversification, considering the possibility of exploiting them as an alternative asset to reduce portfolios' risk. [1] assess investors' reaction to shocks in cryptocurrencies market: they show that there is an asymmetric effect of financial returns on the future volatility of the cryptocurrencies and that the asymmetric effect of returns is reverted so that investors exploit market crashes as an investment opportunity. In [2], an in-depth analysis of Bitcoin volatility and its implications on the diversification effect are described. They conclude that the high level of volatility prevents Bitcoin to perform as a currency. However, [3] show that Bitcoin can be seen as a simpler, less complex version of gold, consistent with the idea that it is digital gold. A more general view was provided by [4], were the relationship between several cryptocurrencies market volatility drivers were analyzed, considering 100 most capitalized cryptocurrencies. All these aspects, together with the assessed conditionality of cryptocurrencies with other macroeconomic variables, lead scientists to

* Corresponding author.
*E-mail address:* gabriella.piscopo@unina.it (G. Piscopo).

https://doi.org/10.1016/j.physa.2025.130770

Available online 24 June 2025

consider fundamental the determination of the correlations of cryptocurrencies with different financial assets to evaluate optimal hedging strategies. Some papers consider the relationship between Bitcoin and commodities to verify the ability of Bitcoin to act as a safe haven against daily movements in commodities: [5] focused on energy and non-energy commodities, applying a bivariate asymmetric pairwise dynamic conditional correlations (DCCs) model. They made use of a general aggregate commodity index, S&P GSCI (which includes energy commodities and non-energy commodities such as agriculture, livestock, industrial metals, and precious metals), the S&P GSCI energy commodities index, and the S&P GSCI non-energy commodities index to test the December 2013 Bitcoin crash. [6] applied Multi-Fractal Detrended Fluctuation Analysis (MF-ADCCA) to analyze the presence and asymmetry of the cross-correlations between the major currency rates (EUR, GBP, YEN, USD), Bitcoin, the Dow Jones Industrial Average (DJIA), gold price (GOLD) and the crude oil market (WTI). [7] considered the time-varying correlations and conditional volatilities between six cryptocurrencies and S&P500 index markets using a copula ADCC-EGARCH model. Some other results may be found in [8–16]. All of the aforementioned works agree that there is an unsubstantial correlation between cryptocurrencies and traditional assets based on historical data, especially during non-crisis periods. This statement implies that cryptocurrencies could be considered as a good diversifier in traditional portfolios, and may be useful in risk management and the anticipation of negative shocks to the market. Recently, [4] show the absence of the leverage effect for cryptocurrencies and the decreasing impact of recent information, suggesting the negligible role of negative volatility. This indicates that negative market shocks have negligible impact on future volatility in the cryptocurrency market. In order to clarify cryptocurrencies behavior, non-linear relations among cryptocurrencies and different commodities were further investigated. [17] considered Bitcoin and Ethereum and applied the methodology of causality on quantiles to deal with the non-linear structure, and examined the predictive power of cryptocurrencies and real-time commodity futures of Gold, Silver, Copper, Crude Oil, Brent Oil, Natural Gas and Wheat on daily basis, spanning the period February 2nd, 2012, to December 31st, 2017. They concluded that significant causal flows are running from cryptocurrency markets to commodity futures, both in terms of mean and volatility. In agreement, [18] examined the nonlinear, asymmetric and quantile effects of aggregate commodity index S&P GSCI and gold prices on the price of Bitcoin using daily data from July 17th, 2010, to February 2nd, 2017, employing several advanced autoregressive distributed lag models (ARDL). They compared the safe-haven roles of Bitcoin, gold (GOLD) and the S&P GSCI against a set of global and country stock market indices (MSCI World, Pimco investment grade bond, US dollar index). It resulted that commodities, including gold, are significant factors driving Bitcoin prices. Importantly, they emphasized the need to apply non-standard models to uncover the complexity and hidden relations between cryptocurrencies and other assets. What emerges from prior studies is that the analyses of the relationship between cryptocurrencies and several economic and financial variables, although widely examined, results unclear, and the need to analyze the potential presence of asymmetric and non-linear effects, which may affect the conclusions, is evident. Moreover, the proposed analyses are conducted considering the daily closing prices of the considered cryptocurrencies or some measures of intraday volatility as in [19].

In this paper, we investigate the relationship between cryptocurrency volatility and the volatility of financial markets by analyzing the relations between the Crypto Volatility Index (CVI), which tracks the 30-day implied volatility of major cryptocurrencies, namely Bitcoin and Ethereum, and six different variables: the Gold volatility index (GVZ), the Crude Oil Volatility Index (OVX), the S&P500 volatility index (VIX), the Dollar to Eur exchange rate (USDEUR), the FED interest rate (FED) and the NASDAQ index. The Crypto Volatility Index has been introduced to assess the growing need of investors for proper estimation of cryptocurrency market volatility, in the same way the S&P 500 volatility index (VIX) is used in stock markets. CVI is a dynamic index, but differently from VIX, it is decentralized and specific to cryptocurrencies. It is also regarded as a "market fear index": it tracks the market's expectations for volatility over the coming 30 days of Bitcoin and Ethereum [20]. The 30-day future fluctuation grade of cryptocurrency market price is obtained using the Black–Scholes option pricing model. Observing the CVI, investors may develop trading strategies for short-term gains and hedge their portfolios against price fluctuations. As regards the volatility indexes used, our choice fell on indices listed on the market in order to understand from real data the cryptocurrency behavior in a global context. In place of the VIX, [21] construct a multivariate volatility index called NetVIX in order to decompose the perturbations into an average volatility and a network amplifier index. They design a synthetic volatility index to catch the network volatility, i.e. the volatility of a basket of different market indexes in a multivariate setting. Differently, we focus our attention on the volatility of a single financial asset, the crypto, and want to study their relation with other financial assets.

In the first instance, we develop an analysis to identify potential non-linear dependencies between variable pairs among CVI and the other selected variables. We consider daily data from January 1st, 2019, to December 31st, 2023 and evaluate and compare three different correlation measures (the Pearson correlation, the distance correlation and mutual information) between pairs of variables to capture general results on the nature of the relationship, which may be linear or non-linear. As a consequence of the observed potential non-linear dependencies between variable pairs, we perform an explanatory analysis of the CVI with the aid of Machine Learning (ML) techniques, in particular, the Random Forest algorithm (RF) and the Gradient Boosting Machines (GBM): these machine learning algorithms thrive in detecting complex interactions and patterns that standard linear models are not able to recognize. In previous contributions, ML techniques were applied to predict cryptocurrency price fluctuations and returns: [22] predicted daily prices of 1681 cryptocurrencies for the period November 2015–April 2018 using three machine learning models, two based on gradient boosting decision trees and one on long short-term memory neural networks. The cited papers focused on the prediction of volatility of cryptocurrency. Recently, [23] present a contribution on the cruptocurrency volatility prediction through generalized autoregressive conditional heteroskedasticity GARCH models, Long Short Term memory (LSTM) network and hybrid models where parameters of GARCH are used as input in the LSTM architecture. The proposed model outperfoms the classical time series volatility forecast models used in the context of cryptocurrency. Our paper offers a different point of view and imply machine learning models shifting the main focus from the crypto volatility forecasts to the analysis of these volatility in a global dynamics. While GARCH models and LSTM networks are inherently designed for time series data and widely used for volatility forecasting,

the machine learning techniques we implement provide a balance between predictive performance and model interpretability, and higher flexibility in handling heterogeneous data without needing stationarity or specific distributional assumptions. Tree-based ensemble methods allow us to derive variable importance measures and provide insights into the driving factors of crypto volatility. [24] analyzed the most liquid twelve cryptocurrencies at the daily and minute level frequencies from April 1st, 2013 to June 23rd, 2018 using ML classification algorithms including the support vector machines, logistic regression, artificial neural networks, and random forests. [19] developed a suitable model to capture the cryptocurrency volatility dynamics for the period April 28th, 2013 to December 15th, 2019 based on Jordan Neural Network using data from Bitcoin, Ripple, and Ethereum. Other results may be found in [25–29] and references therein.

In this paper, we apply RF and GBM to investigate the relationship between the CVI and six explanatory variables. Our numerical results show consistent predictive performance, as evidenced by the alignment between the observed and predicted density functions of the CVI, supporting the reliability of the explanatory analysis. In particular, the RF algorithm provides more accurate outcomes and avoids overfitting. Furthermore, we evaluate the weighted impurity measure [30] to analyze the magnitude of each input variable's effect on predictions: outcomes indicate that the dependent variable CVI is mostly influenced by financial returns, while the impact of variables expressing financial volatility results to be low. Therefore, we capture the marginal effects for the most impactful variables through the Partial Dependence Plot (PDP), analyzing ML algorithms' performances. The overall interaction strength is then obtained by evaluating the H statistic as in [31]. In order to complete our analysis and compensate for possible "black boxes" due to ML techniques usage, we pursue a SHAP analysis on three different values of the CVI to show the contribution of each considered variable on the CVI explanation. Our conclusions point out the need to explore broader analysis that also takes non-linear relationships into account since, when focusing on linear interaction, some aspects and influential features may remain undetected.

The paper is structured as follows: Section 2 contains a resume of machine learning techniques, with a focus on Random Forest and Gradient Boosting Machines. Section 3 contains data description and the investigation of dependence between the considered variables. Section 4 includes the applications and the results of ML techniques. The paper is concluded by Section 5.

## 2. Machine learning techniques: Random forest and gradient boosting machines

Suppose that $x = (x_1, \dots, x_N)$ denotes the vector of input variables and $y$ represents the response (or target) variable. The problem is to estimate the relationship between $y$ and $x$, i.e. we need to construct the functional $f$ which describes the dependence between $x$ and $y$. Given the relation

$$y = f(x) + \epsilon,$$

the machine learning algorithms may help construct an approximation $\hat{f}(x)$ in a way that minimizes the reducible error $E[f(x) - \hat{f}(x)]^2$. Among the machine learning algorithms, we will consider the Random Forest and the Gradient Boosting Machine.

The Random Forest algorithm (RF) is a versatile tool that can be successfully applied to classification and regression tasks. It is based on an ensemble learning mechanism: the accuracy of the prediction is improved by leveraging the combination of multiple models. RF, introduced in [32] as an extension of the bagging procedure [33], allows for a random split selection of the variables in the input dataset, i.e. it introduces a randomization procedure and arbitrarily generates a series of decision trees to be combined to predict the correct output. Therefore, differently from bagging, RF searches for the best split among a random subset of variables for each node, instead of looking for the best split among all input variables. Then, final predictions from each tree are combined via aggregation techniques, which in the case of correlations is usually the average operator. The main benefits of applying the Random Forest algorithm are the reduction of the risk of overfitting and training time and the high level of accuracy and efficiency on large datasets. Moreover, RF uses an Out-of-Bag (OOB) estimate for error assessment. It produces a built-in validation process without the need to consider a separate validation dataset. It is indeed [32] in his original article where he proposed the RM who emphasizes that 'unlike cross-validation, where bias is present but its extent unknown, the out-of-bag estimates are unbiased'. In details, the RF estimator $\hat{f}(x)$ is calculated as follows:

$$\hat{f}(x) = \frac{1}{N} \sum_{n=1}^{N} \hat{f}^{(n)}(x),$$

where $N$ indicates the number of decision trees obtained from the bootstrap of the original training dataset. We make use of the R package `randomForest` to construct the RF estimator [34]. To run the algorithm, some of the parameters one needs to specify are the number of trees to grow (*ntree*), the minimum size of terminal nodes (*nodesize*) and the maximum number of terminal nodes (*maxnodes*). The number of trees should be accurately set to ensure a good prediction and not ignore some input rows. In the Random Forest algorithm, each tree is trained on a randomly drawn sample of the data (bootstrap sampling), and at each split, only a random subset of features is considered (feature randomization). Consequently, key parameters such as the number of trees, the depth of each tree, and the number of selected features must be carefully tuned. In particular, increasing the number of trees generally reduces variance and improves accuracy, as the algorithm averages predictions across multiple trees. Each tree provides a different representation of the data, and having more trees enhances the diversity of input selections. Moreover, averaging their predictions helps mitigate variance. [35] investigated whether there is an optimal number of trees and found that, beyond a certain threshold, increasing the number of trees yields negligible accuracy improvements while significantly increasing computational costs. The number of trees should be chosen in conjunction with the number of features each tree can uses. A smaller number of selected features may lead to higher bias in individual trees. Since both variance and bias contribute to the overall estimation error,

reducing variance by increasing the number of trees can be a reasonable approach. However, the effect of the number of features on model performance is complex. Generally, increasing the number of available features improves accuracy, as the algorithm has more options to select from at each node. However, this also reduces diversity among trees and accelerates the training process. The same attention must be paid to the minimum size of terminal nodes and the maximum number of terminal nodes that have to be adequate to ensure the right dimension and the right number of trees to be grown.

The Gradient Boosting Machine (GBM) is another ensemble machine learning algorithm that combines multiple weak predictors to generate the final model. In particular, the GBM sequentially minimizes a loss function via a gradient descent procedure and trains the new models by placing more weights on instances with erroneous predictions. In this way, the prediction errors are reduced and the model's accuracy is enhanced. The possibility to choose the most appropriate loss function makes this algorithm highly flexible and customizable [31]. The GBM algorithm operates by estimating the approximation $\hat{f}(x)$ in a way that minimizes the expected value of the specified differentiable loss function $\Phi(y, f)$. The problem can be written in the following way:

$$\hat{f}(x) = \arg\min_{f(x)} E_x[E_y \Phi(y, f(x))|x].$$

In the iteration procedure, at each step $i$, the algorithm "boosts" the approximation adding an estimator $h_i$ to provide a better model:

$$\hat{f}_i(x) = \hat{f}_{i-1}(x) + \lambda_i h_i(x),$$

where $h_i$ is a base learner differentiable function and $\lambda_i$ is a multiplier obtained by solving the optimization problem. For each iteration, a different minimization problem is solved and a different multiplier is selected. In this work, we make use of the R package `gbm` to implement gbm as in [31] and find the GBM estimator $\hat{f}(x)$. The `gbm` package requires choosing the number of trees (*n.trees*), which represents the number of GBM iterations: this value has to be accurately assigned since a high value reduces the error on the training set. With the same number of trees, the depth of each of them, i.e. the maximum number of node or split, can decrease or increase. Increasing the depth can generate overfitting. In fact, deeper trees can capture more complex relationships in the data, but increasing it too much might catch noice or a complex structure that is specific of a particular randomly selected subset and this can generate overfitting. Instead, increasing the number of trees reduces the variance of the estimator because it averages over a greater number of trees. For this reason, as the number of trees increases, the error is reduced. Other key parameters are represented by the number of cross-validation folds (*cv.folds*), which is chosen according to the dataset (usually 20% of the data is considered sufficient to test the model), the depth of each tree involved in the estimate (*interaction.depth*) and the learning rate parameter (*shrinkage*), which controls the size of incremental steps during the iteration. Note that complex models with rich tree structure, i.e. with assigned high value for *n.trees* and *interaction.depth*, may provide almost no advantage over compact trees.

## 3. Numerical application

### 3.1. Data description

The variable under investigation is the Crypto Volatility Index (CVI). Our objective is to explore the relationship between the CVI and the volatility of traditional financial markets, specifically examining the Gold volatility index (GVZ), Crude Oil volatility index (OVX) and S&P500 volatility index (VIX). Moreover, we consider three other explainable variables, the Dollar to Eur exchange rate (USDEUR), the FED interest rate (FED) and the NASDAQ index. Eq. (1) describes the proposed machine learning model:

$$CVI_t \sim GVZ_t + OVX_t + VIX_t + USDEUR_t + FED_t + NASDAQ_t. \tag{1}$$

CVI is a dynamic index used as a benchmark for tracking cryptocurrency volatility by analyzing the price volatility of cryptocurrency options. CVI is calculated on the 30-day implied volatility of Bitcoin and Ethereum, applying the Black & Scholes formula to the respective options. The CVI index is measured on a scale of 0 to 200, where the higher the value of CVI, the greater the cryptocurrency volatility.

Gold can be considered a haven in times of economic and political uncertainty and typically exhibits different volatility than the stock market. The Cboe Gold ETF Volatility Index (GVZ) estimates the expected volatility of 30-day returns of a physically backed gold exchange-traded fund, the SPDR Gold Shares ETF (GLD), the largest gold exchange-traded fund. Like the other volatility indexes considered, the GVZ offers an estimate of the volatility of gold by considering the volatility of options on the underlying GLD. Specifically, the GLZ is calculated by interpolating two time-weighted sums of option price values over 30 days. The GLZ takes into account the premiums that traders are paying for options on the GLD and uses them to determine what prices the underlying should reach before each option expires. This provides information on the market's expectations of gold price movements in the coming month.

The Cboe Crude Oil ETF Volatility IndexSM (OVX) is an estimate of the expected 30-day volatility of crude oil. Like the GVZ, it is calculated by considering the volatility of the options whose underlying is the United States Oil Fund (USO ETF).

The Cboe Volatility Index (VIX) is a dynamic index that represents the market's expectations on the price changes of the S&P500 index (SPX). In the same manner, as the previous indexes, it generates a 30-day forward projection of volatility working on the price of options whose underlying is the S&P500 index.

We collected daily data from January 1st, 2019 to December 31st, 2023. Data has been retrieved from Yahoo Finance and Investing.com. In Table 1, we report the summary statistics of the data set, which provide a clear picture of the distribution and variability for each variable, highlighting the most volatile and stable markets.

**Table 1**
Summary statistics.

| Variable | Mean | StdDev | Min | Max | Q1 (25%) | Median | Q3 (75%) |
|---|---|---|---|---|---|---|---|
| CVI | 79.29 | 21.72 | 35.89 | 170.55 | 63.75 | 76.78 | 93.23 |
| GVZ | 17.08 | 4.61 | 8.88 | 48.98 | 14.76 | 16.47 | 19.04 |
| OVX | 45.25 | 24.30 | 24.00 | 325.15 | 35.02 | 39.36 | 47.49 |
| VIX | 22.17 | 7.21 | 11.71 | 72.63 | 17.05 | 20.77 | 25.88 |
| USDEUR | 0.90 | 0.05 | 0.81 | 1.05 | 0.86 | 0.90 | 0.93 |
| FED | 1.83 | 1.94 | 0.04 | 5.33 | 0.08 | 1.55 | 3.08 |
| NASDAQ | 6883.16 | 1470.71 | 4030.77 | 9855.42 | 5723.81 | 6806.19 | 8101.50 |

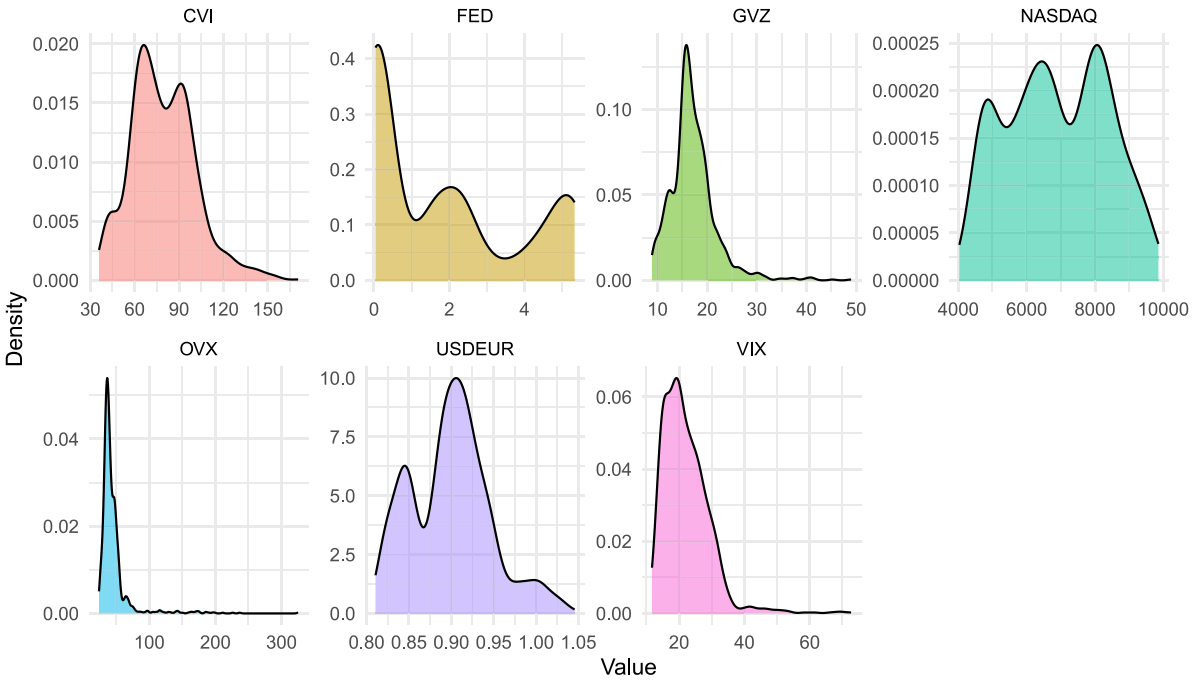



**Fig. 1.** Density functions of the variables in the model.

The CVI shows a high average level of volatility, with a mean of 79.29 and a standard deviation of 21.72, indicating significant variability. Its range extended from 35.89 to 170.55, with the median value (76.78) closer to the first quartile (63.75), suggesting a slightly skewed distribution toward higher values. Differently, the GVZ displayed lower and more stable levels of volatility, with a mean of 17.08 and a standard deviation of 4.61. The range of GVZ values (8.88 to 48.98) highlights occasional periods of elevated volatility, while the balanced distribution is reflected in the median value of 16.47.

The OVX had the highest variability among the indices, with a mean of 45.25 and a standard deviation of 24.30. Its range (24.00 to 325.15) indicates episodes of extreme volatility, while the relatively low median (39.36) and first quartile (35.02) suggest that these high values are outliers. Conversely, the VIX demonstrated moderate volatility, with a mean of 22.17 and a standard deviation of 7.21. Its range (11.71 to 72.63) and the median value (20.77), close to the first quartile (17.05), suggest a distribution slightly skewed toward higher values.

The USDEUR exchange rate showed remarkable stability, with a mean of 0.90, a low standard deviation of 0.05, and a range of 0.81 to 1.05. The proximity of the quartiles and median indicates an almost uniform distribution. The FED displayed moderate variability, with a mean of 1.83 and a standard deviation of 1.94. Its range (0.04 to 5.33) reflects economic conditions during the observation period, with lower rates dominating, as evidenced by the first quartile (0.08) and median (1.55).

Finally, the NASDAQ index showed high performance and variability, with a mean of 6883.16 and a standard deviation of 1470.71. Its range (4030.77 to 9855.42) underscores substantial market movements, while the median value (6806.19), slightly below the mean, suggests a distribution somewhat concentrated toward lower values.

The variables' density functions are depicted in Fig. 1, where each plot corresponds to the distribution of one variable in the dataset.

Even if the variables considered are time series, the RF algorithm learns complex relationships between input and output but has no awareness of time and treats each row of the dataset independently. This represents an important limitation in forecasting, as failing to account for time dependence, the non-stationarity of the series, and correlations between variables can lead to distorted

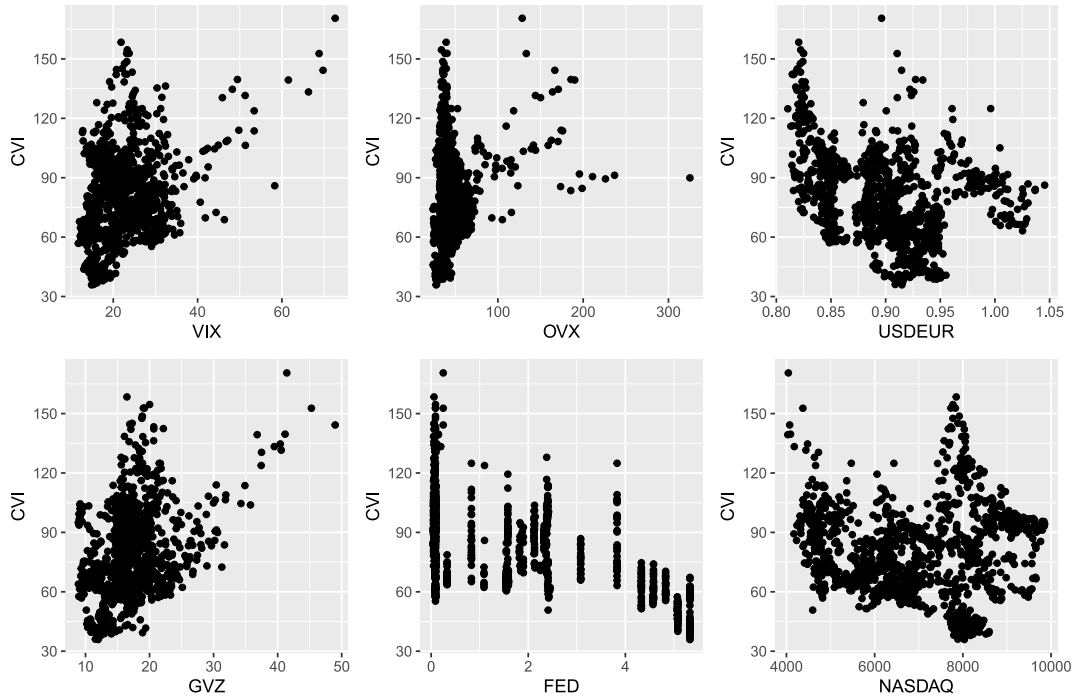



**Fig. 2.** Scatter plots of the CVI compared to the other variables in the model.

analyses and misleading results. Since our objective is explanatory rather than predictive, the analysis remains valuable, despite the acknowledged limitations. By interpreting the results critically, we can extract meaningful insights while maintaining awareness of potential causal gaps. This approach ensures that, even without capturing temporal dependencies, the findings contribute to a deeper understanding of the relationships within the data.

### *3.2. Investigation of dependence between variables*

It is well known that machine learning techniques are powerful non-linear regression algorithms designed to model and predict complex, non-linear relationships between input and target variables. These algorithms excel in capturing intricate patterns and interactions that linear models cannot effectively represent. To explore whether such relationships exist within our dataset, we develop an analysis to identify potential non-linear dependencies between variable pairs. This would provide a strong justification for employing machine learning algorithms to more effectively investigate and model these relationships.

Firstly, we use the scatter plot (Fig. 2) to visualize the potential non-linear relationship between CVI and the other variables in the dataset.

Secondly, we conduct a correlation analysis, beginning with the traditional Pearson correlation matrix, followed by the distance correlation and mutual information between pairs of variables.

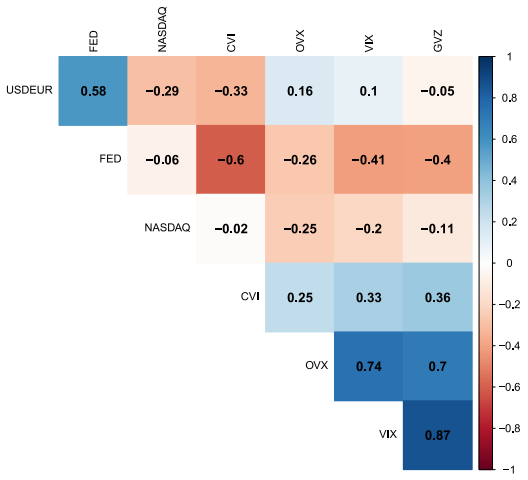
Unlike Pearson correlation, which is limited to capturing only linear relationships, distance correlation and mutual information can quantify the dependency between variables in a more general way, regardless of the nature of the relationship (linear or non-linear). If distance correlation or mutual information is high but Pearson correlation is low, it indicates a non-linear relationship between the variables.

Distance correlation helps identify non-linear relationships between variables. A high distance correlation suggests a relationship, which may be linear or non-linear. Distance correlation values range from 0 to 1. A value of 0 indicates no dependency, while a value closer to 1 signifies a strong dependency.

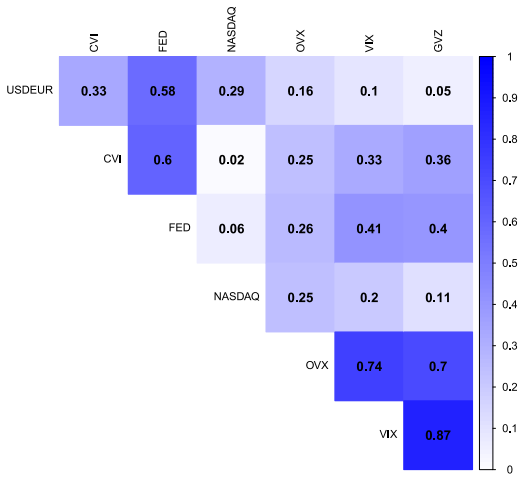
Mutual information measures the amount of information one random variable contains about another. Each element in the mutual information matrix represents the mutual information between two variables. Mutual information values range from 0 to 1. A value of 0 indicates no mutual information (i.e., no dependence). Higher values, closer to 1, indicate stronger mutual dependence.

Since the scale of the values for Pearson correlation differs from that of distance correlation and mutual information, we take the absolute value of the Pearson correlation matrix. This allows us to compare the strength of relationships between variables across the different dependency measures.

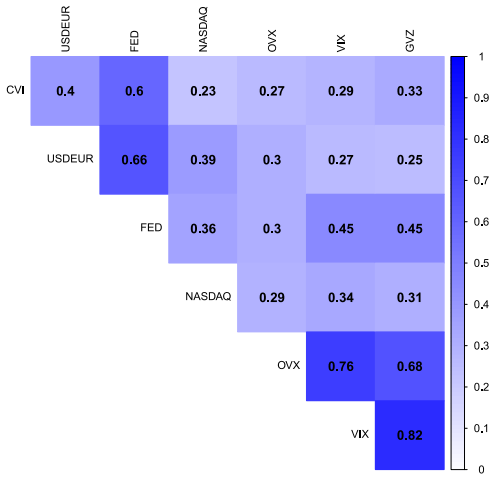
Fig. 3 shows the Pearson correlation matrix (top left panel), the Pearson correlation matrix in absolute values (top right panel), the distance correlation matrix (bottom left panel), and the mutual information matrix (bottom right panel).



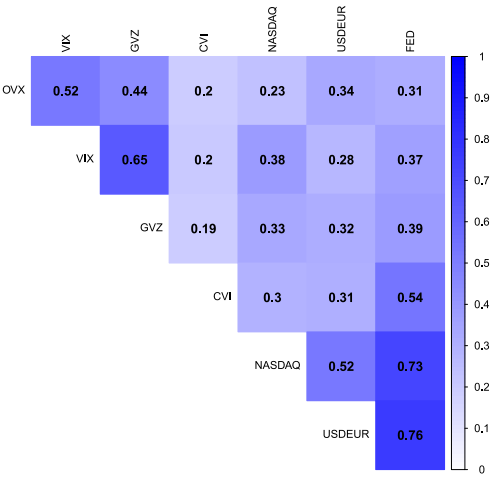

(a) Pearson correlation matrix




(b) Pearson correlation matrix (abs)




(c) Distance correlation matrix




(d) Mutual information matrix.

**Fig. 3.** Pearson correlation matrix (values (a), and absolute values (b)), Distance correlation matrix (c), Mutual information matrix (d).

From Fig. 3, panel (a), we note that CVI has a linear positive correlation with the other volatility indexes (VIX, GVZ, OVX) and a negative correlation with FED, NASDAQ and USDEUR. In other words, if we analyze the linear correlation, we observe that when the financial and raw materials markets are more volatile, cryptocurrency trading also presents greater volatility, although with opposite movements compared to those recorded by the financial returns. This information could support the use of cryptocurrency for portfolio diversification, in terms of diversification of returns, although the fund manager must pay attention to the overall riskiness. However, the information on the linear relationship is not sufficient to grasp the dependencies between cryptocurrency and financial markets and our analysis continues with the following in-depth analysis.

We compute the Mean Absolute Error (MAE) of the differences between each pair of correlation matrices to summarize their overall difference. The MAE for the difference between Pearson correlations and distance correlations (mutual information) is 0.08 (0.15). The results indicate more relevant differences between the Pearson correlation and mutual information than between the Pearson correlation and distance correlation.

To visualize where the discrepancies occur, we create heatmaps for the pairwise difference in absolute values (Fig. 4). A gradient color scale represents the differences between the two types of correlation measures. Lower differences are in light blue, and higher differences are in red, with intense blue for zero difference. The differences between Pearson correlations and distance

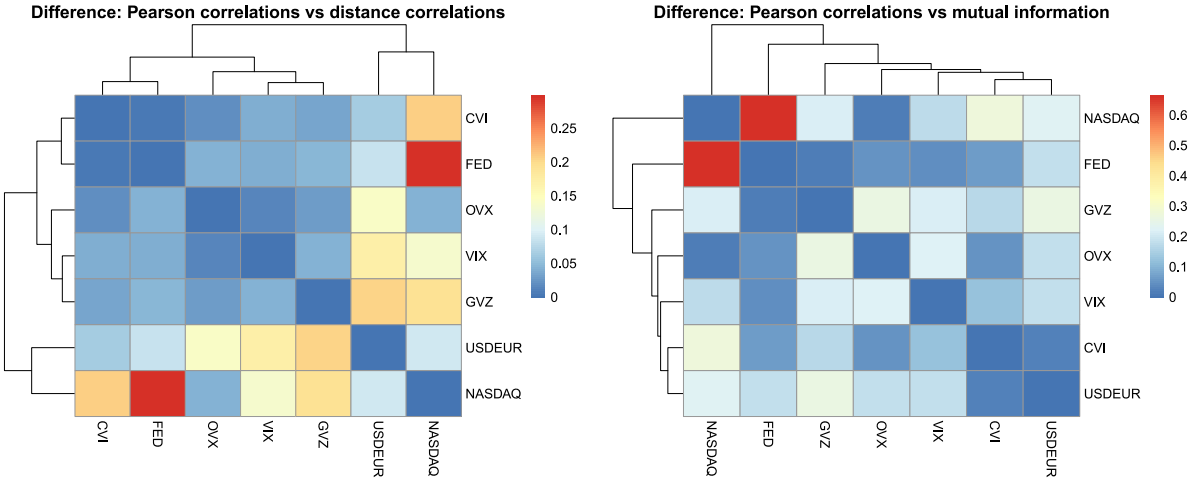



**Fig. 4.** Heatmaps for the pairwise difference: in the left panel is represented Pearson correlations vs distance correlations, while in the right panel Pearson correlations vs mutual information.

correlations range from approximately 0 to 0.3, while the differences between Pearson correlations and mutual information range from approximately 0 to 0.7.

Based on the difference between Pearson correlations and distance correlations, non-linear relationships are observed primarily between FED and NASDAQ (0.30), CVI and NASDAQ (0.21), and GVZ and USDEUR (0.20). If we had looked only at the linear dependence, we would have concluded a poor dependence between CVI and NASDAQ, but the results of the non-linear dependence led us to a different conclusion. From the difference between the Pearson correlations and mutual information, we find non-linear relationships primarily between FED and NASDAQ (0.67), CVI and NASDAQ (0.30), GVZ and USDEUR (0.26), and GVZ and OVX (0.26). These values indicate that traditional Pearson correlation may miss key patterns in these variable interactions, which distance correlation and mutual information can capture.

To directly compare and visualize values between matrices, we also show the pairwise scatter plots. The scatter plot in the left panel (Pearson vs. Distance Correlation) shows a more concentrated range of values (e.g., 0 to 0.3), while the scatter plot in the right panel (Pearson vs. Mutual Information) shows a wider range of values (e.g., 0 to 0.7). This could indicate that mutual information captures a broader spectrum of dependencies, including nonlinear ones that Pearson might not detect. This suggests that mutual information is more sensitive to a wider variety of relationships. The points on the scatter plot in the left panel (Pearson vs. Distance Correlation) are more tightly clustered around a diagonal line (indicating a consistent relationship), suggesting that the Pearson and distance correlation methods are relatively aligned in how they capture relationships between the variables. The points on the scatter plot in the right panel (Pearson vs. Mutual Information) appear more dispersed, with a wider spread along the $x$-axis and $y$-axis, suggesting that mutual information detects more fine dependencies (perhaps nonlinear relationships or interactions), which Pearson does not capture as effectively (see Fig. 5).

In summary, from the analysis of the dependence between variables, we conclude that there is a non-linear relationship between some pairs of variables in our dataset.

## 4. Machine learning results and their comparison with Ordinary Least Squares (OLS)

We present the machine learning results and compare them to those from an Ordinary Least Squares (OLS) regression.

Machine learning algorithms depend on various parameters, some of which are fine-tuned through an optimization process. After splitting the dataset into a training set (80% of the data) and a test set (the remaining 20% of the data), we conduct a tuning procedure of the models' parameters on the training set employing a grid search strategy, which systematically examines a predefined set of parameter values to determine the optimal configuration.

In the RF algorithm, the tuning parameters considered are the number of predictors considered at each split of the tree (`mtry`), the number of trees (`ntree`) and the minimum size of terminal nodes (`nodesize`), while in the GBM, the number of boosting iterations (`ntree`), the learning rate (`shrinkage`), which controls the rate at which algorithm learns, the interaction depth (`interaction.depth`), which is the maximum depth of the trees, and the minimum number of observations in the terminal node (`n.minobsinnode`) controlling the complexity of the algorithm.

Table 2 outlines the machine learning algorithms we use, their corresponding R-packages, the tuning parameters, their ranges, and the resulting optimal values.

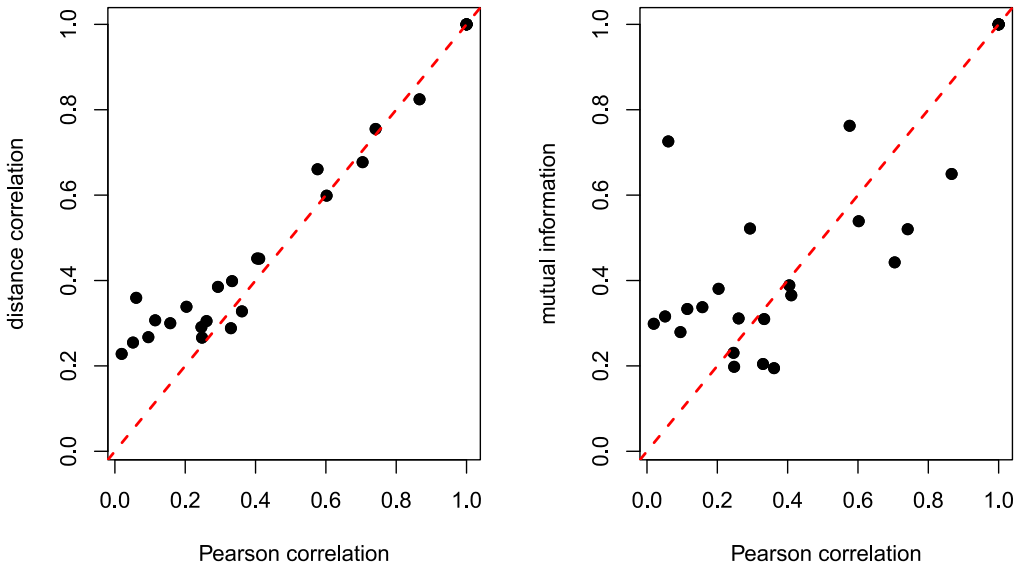



**Fig. 5.** Pairwise scatter plots: in the left panel is represented Pearson correlations (abs) vs distance correlations, while in the right panel Pearson correlations (abs) vs mutual information.

**Table 2**
Machine learning algorithms and hyperparameters used for tuning.

| Algorithm | R-package | Tuning parameter | Range | Optimal value |
|---|---|---|---|---|
| RF | randomForest | `ntree` | [200, 500] step 100 | 200 |
| | | `mtry` | [2, 5] step 1 | 3 |
| | | `nodesize` | [3, 5, 10] | 5 |
| GBM | caret | `n.trees` | [1000, 2000, 3000] | 3000 |
| | | `shrinkage` | [0.01, 0.05, 0.1] | 0.01 |
| | | `interaction.depth` | [2, 3, 5] | 5 |
| | | `n.minobsinnode` | [5, 10] | 10 |

**Table 3**
OLS regression results.

| Variable | Estimate | Std. Error | t value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 9.500e+01 | 1.572e+01 | 6.045 | 2.15e−09*** |
| VIX | −1.854e−01 | 1.836e−01 | −1.010 | 0.31281 |
| OVX | 7.561e−03 | 4.093e−02 | 0.185 | 0.85348 |
| USDEUR | −1.092e+01 | 1.735e+01 | −0.629 | 0.52931 |
| GVZ | 8.630e−01 | 2.634e−01 | 3.277 | 0.00109** |
| FED | −6.054e+00 | 4.304e−01 | −14.067 | < 2e-16*** |
| NASDAQ | −8.254e−04 | 4.093e−04 | −2.016 | 0.04404* |

* Significance code: $p < 0.05$
** Significance code: $p < 0.01$
*** Significance code: $p < 0.001$

*Ordinary Least Squares (OLS) results*

In Table 3, we report the results of an OLS regression, displaying the estimated coefficients, standard errors, t-values, and significance levels for the independent variables.

The intercept suggests a strong baseline effect, while GVZ has a significant positive influence on the dependent variable. In contrast, FED shows a strong negative effect, indicating substantial impact. NASDAQ shows a small but statistically meaningful negative relationship. Meanwhile, VIX, OVX, and USDEUR have limited influence in this model.

The predictive performances of each model is first assessed using the density functions of the observed values compared to those predicted by RF, GBM, and OLS, for both the training dataset and the testing dataset (see Fig. 6). The density curve of observed values (in black) is compared against RF (red), GBM (blue), and OLS (green) predictions. The graph referring to training data demonstrates how closely the RF and GBM predictions align with the actual observed data. In contrast, the OLS model fails to capture the true distribution, as evidenced by its density curve deviating significantly from the observed data. The OLS predictions

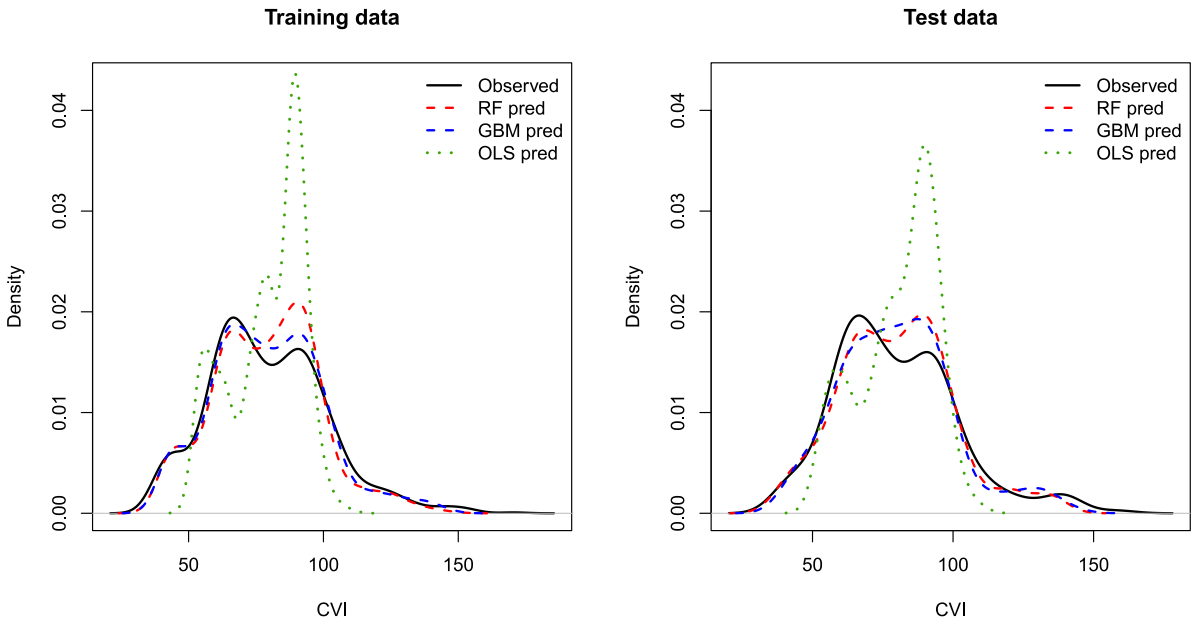



**Fig. 6.** Density functions of the CVI observed values compared to those predicted by RF, GBM, and OLS.

**Table 4**
Performance metrics for RF, GBM, and OLS models.

| Model | Train | | | Test | | |
|---|---|---|---|---|---|---|
| | $R^2$ | RMSE | MAPE | $R^2$ | RMSE | MAPE |
| RF | 0.9084 | 6.689 | 5.866 | 0.9047 | 6.713 | 5.657 |
| GBM | 0.9505 | 4.895 | 4.771 | 0.8744 | 7.669 | 6.843 |
| OLS | 0.3820 | 17.087 | 17.332 | 0.3833 | 16.991 | 16.673 |

appear more dispersed and do not reproduce the peaks present in the actual density function. This indicates that the OLS model struggles to represent the underlying data structure, likely due to its linear assumptions, which may be inadequate for capturing the more complex, non-linear relationships in the dataset.

Like the training data graph, the test data graph displays the density curves for RF and GBM predictions compared to the observed values, providing insight into the models' performance on unseen data. Since the test data predictions appear reasonably aligned with the observed densities, the models seem to generalize well to unseen data, making overfitting less likely. Notably, the predictions remain consistent between the training and test datasets.

In contrast, the OLS model once again fails to capture the true distribution, with its density curve showing a poor fit to the observed values.

About the difference between the two ML algorithms, RF (red curve) appears to provide a closer fit to the observed density, especially in regions with higher values, suggesting its effectiveness in capturing extreme data points. GBM (blue curve) shows a comparable fit but seems to focus more accurately on the central tendencies (expected values).

An alternative visualization of the results is depicted in Fig. 7, where the $x$-axis represents the number of observations, and the $y$-axis represents the CVI values of the observations. The closeness of the RF and GBM predicted lines (red and blue, respectively) to the observed values (black points) indicates the accuracy of each model. Overall, we can appreciate the good performance of both these models on the training and test datasets. The OLS predictions (green line) appear to deviate more from the observed values compared to RF and GBM, particularly in capturing the variability of the data. The greater dispersion of the OLS line suggests that it is unable to reflect the underlying structure of the observations.

We complement this visual assessment with three quantitative metrics, R-squared and standard error measures, root mean square error (RMSE) and mean absolute percentage error (MAPE), for both datasets to confirm whether the model's test performance drops significantly compared to its training performance.

For RF model, the $R^2$, RMSE, and MAPE values remain very close between the train and test sets, suggesting that the model generalizes well without significant overfitting. For the GBM model, we observe a noticeable drop in $R^2$ (from 0.9505 to 0.8744) and an increase in RMSE and MAPE in the test set, indicating that the model may have been slightly overfit to the training data. In the case of the OLS model, the similarity of the performance metrics between training set ($R^2 = 0.3820$; RMSE = 17.087; MAPE = 17.332) and test set ($R^2 = 0.3833$; RMSE = 16.991; MAPE = 16.673) indicates that overfitting is not an issue. However, its overall

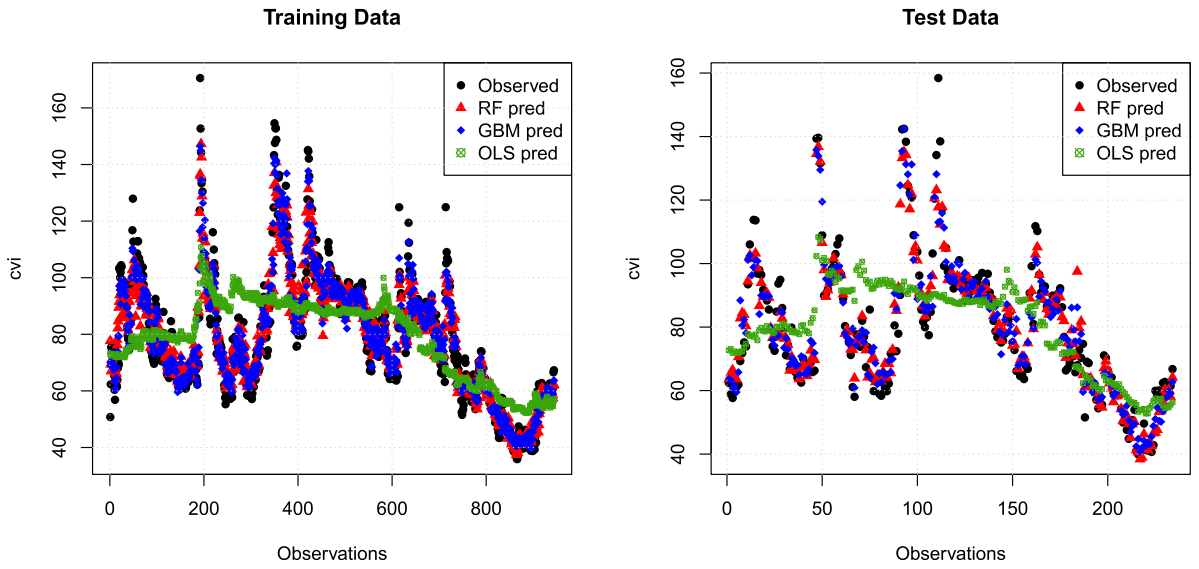



**Fig. 7.** Results: observed versus predicted values by RF, GBM, and OLS.

**Table 5**
RFE results. RF.

| Variables | RMSE | $R^2$ | MAE | RMSESD | $R^2$SD | MAESD |
|---|---|---|---|---|---|---|
| 1 | 13.945 | 0.5878 | 10.011 | 1.671 | 0.0817 | 0.956 |
| 2 | 11.286 | 0.7310 | 7.963 | 1.515 | 0.0628 | 0.836 |
| 3 | 8.821 | 0.8377 | 6.086 | 1.118 | 0.0379 | 0.673 |
| 4 | 7.942 | 0.8715 | 5.548 | 1.098 | 0.0358 | 0.620 |
| 5 | 7.649 | 0.8835 | 5.379 | 1.054 | 0.0274 | 0.598 |
| 6 | 7.140 | 0.8962 | 4.987 | 1.028 | 0.0280 | 0.610 |

accuracy remains low, suggesting that the underlying relationships between predictors and CVI are likely nonlinear and involve interactions that the linear specification of OLS fails to capture.

To assess the accuracy of the RF and GBM competing predictions, we apply the Diebold–Mariano (DM) test [36]. Given the prediction error $e_i$, defined as the difference between the predicted values from model $i$ and the actual values, the loss associated with the prediction from model $i$ is assumed to be $g(e_i) = e_i^2$. The two predictions are considered equally accurate if and only if the loss differential between them, defined as $d = g(e_1) - g(e_2)$, has zero mean.

The DM test statistic is given by $DM = \frac{\bar{d}}{\sqrt{\frac{s}{N}}}$, where $\bar{d}$ is the sample mean of the loss differential, $s$ is its variance, and $N$ is the sample size. The test's null hypothesis states that both models have the same predictive accuracy, i.e., $H_0 : E[d] = 0$. We set the alternative hypothesis as follows: $H_1 : E[d] < 0, \quad \forall t$, which implies that the expected loss of method 1 is lower than that of method 2, meaning method 1 provides more accurate forecasts than method 2. Under $H_0$, the DM statistic follows an asymptotic standard normal distribution with a mean of 0 and a standard deviation of 1.

The test statistic DM is −3.1359, indicating that the GBM model is less accurate than the RF. The $p$-value is 0.0009669, below the significance level of 0.05, implying rejection of the null hypothesis.

In order to investigate the impact of using fewer variables on model performance, we have implemented a popular feature selection method, the Recursive Feature Elimination (RFE). This method iteratively eliminates features and evaluates the model's performance to identify the optimal subset of variables. In the RF model, the RFE algorithm finds that all the variables are important, emphasizing that each variable contributes meaningfully to the model, albeit to varying degrees (in order of importance: FED, USDEUR, NASDAQ, GVZ, OVX, and VIX). In the GBM, the RFE algorithm finds that all the variables are important except VIX. In the following tables, we show the resampling performance over the subset size for RF (Table 5) and GBM (Table 6).

Given these results, we again implement the GBM, excluding VIX from the model, and compare its performance to that of RF, including all variables. The hyperparameters' tuning of GBM has consequently been updated (see Table 7).

Furthermore, we apply the DM test to compare the new GBM with the original RF. The test statistic DM is −3.1178, again indicating that the GBM model is less accurate than the RF. The $p$-value is 0.001026, below the significance level of 0.05, implying rejection of the null hypothesis.

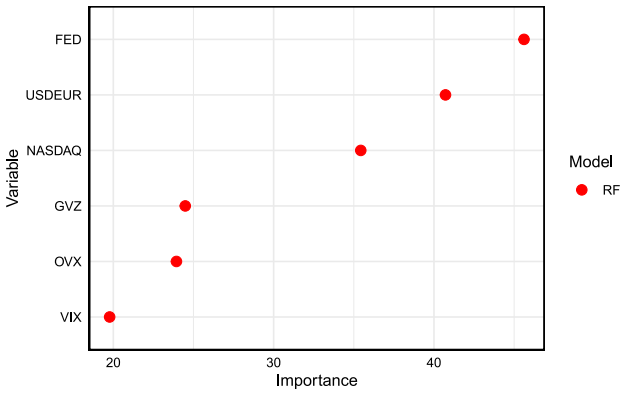
Having verified the superiority of the RF model over the GBM, we will henceforth focus exclusively on the former.

**Table 6**
RFE results. GBM.

| Variables | RMSE | $R^2$ | MAE | RMSESD | $R^2$SD | MAESD |
|---|---|---|---|---|---|---|
| 1 | 14.303 | 0.5683 | 10.349 | 1.3561 | 0.04285 | 0.5628 |
| 2 | 12.508 | 0.6705 | 9.273 | 0.5450 | 0.02630 | 0.3307 |
| 3 | 10.629 | 0.7622 | 7.733 | 0.7020 | 0.03403 | 0.5318 |
| 4 | 10.009 | 0.7882 | 7.251 | 0.8832 | 0.04771 | 0.6461 |
| 5 | 9.813 | 0.7971 | 7.200 | 0.5909 | 0.02442 | 0.3895 |
| 6 | 9.863 | 0.7945 | 7.172 | 0.7377 | 0.02795 | 0.4549 |

**Table 7**
Training and testing performance for GBM (excluding VIX)

| Model | Train | | | Test | | |
|---|---|---|---|---|---|---|
| | $R^2$ | RMSE | MAPE | $R^2$ | RMSE | MAPE |
| GBM | 0.9467 | 5.073 | 4.983 | 0.8745 | 7.666 | 6.811 |




(a) RF

**Fig. 8.** Variable importance plots.

### 4.1. Explainability tools

*Variable importance.* We investigate the magnitude of each input variable's effect on predictions through the feature importance measure. Fig. 8 shows how the features are influential in predicting the dependent variable CVI. The plot is based on the weighted impurity measure introduced by [30]. It is calculated by summing up the weighted impurity Gini index for all tree nodes where a given input variable is used and averaged over all the trees in the forest. We can note that the variables that express financial volatility are less influential than those expressing financial returns. So cryptocurrency volatility is more related to financial returns than financial volatility itself. In other words, the perceived riskiness of the financial markets has less impact on the volatility of cryptocurrency trading than the returns. Another aspect to highlight is that if we look only at the linear relationship, the impact of stock returns measured by the NASDAQ index is negligible, while in a broader analysis that also takes non-linear relationships into account, the NASDAQ is among the top three most influential variables.

*Partial dependence plot.* Fig. 9 illustrates the Partial Dependence Plot (PDP) outputs, showing the relationship between the three most important input variables and the model's predictions of the target variable, CVI. It represents the marginal effect of one or more features on the prediction of the CVI. To understand how the plot is constructed, we refer to [31]. Let us consider the feature of interest $X_S$, and the set $C$ of all the other features $X_C$. The partial dependence (PD) function on $X_S$ (say $f_S^{PD}(x_S)$) is the marginal expectation over the features in the set $C$:

$$f_S^{PD}(x_S) = \mathbb{E}_{X_C}\left[f(x_S, X_C)\right].$$

The PD function of $f$ on $X_S$ can be written as:

$$f_S^{PD}(x_S) = \int f(x_S, X_C) d\,P(X_C)$$

and can be easily estimated on the observed data. In Fig. 9, the $x$-axis shows the values of the variable being analyzed (FED in red, NASDAQ in green, and USDEUR in blue). The $y$-axis represents the partial dependence, illustrating the average model prediction as

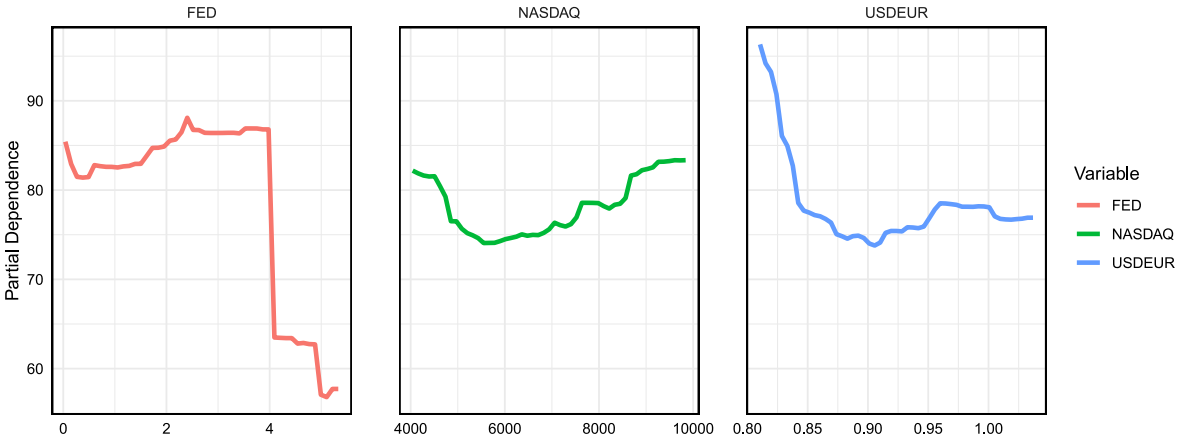



**Fig. 9.** Partial Dependence Plots (PDP) for the most important predictors. RF model.

the variable's value changes, holding other variables constant. Each variable's PDP is plotted independently to display how changes in its value influence the ML algorithm's predictions.

The PDPs reveal interesting insights into how the RF algorithm captures the relationships between variables and predictions. When examining FED, the curve provided by the model shows short changes, reflecting its ability to adapt more effectively to localized shifts in the data. For NASDAQ, the model maintains a steady and systematic response across the range, capturing broader trends in the data. For USDEUR, the model shows a smooth curve, suggesting that it captures the relationship in a non-linear and continuous manner.

Overall, the PDP in Fig. 9 suggests that changes in the FED, NASDAQ, and USDEUR variables are associated with a nonlinear response in the CVI.

For the FED variable's PDP, there is an increase in the CVI within the FED value range of 1 to 4, followed by a sharp drop after 4, where the CVI declines from 87 to 69. This is then followed by periods of stability or slower changes, with values stabilizing around 60.

The PDP for NASDAQ indicates a positive relationship, with higher NASDAQ values corresponding to higher CVI levels. This suggests that the CVI tends to increase as NASDAQ performance rises, which could reflect a correlation between equity markets and crypto markets. For the USDEUR exchange rate, the PDP exhibits a distinct pattern, characterized by fluctuations across its range of values.

Finally, once we catch the marginal effects for the most important features through the PDP, we can consider the strength of the overall interaction of each single feature with the others.

*Feature interaction.* The feature interaction calculates how much of the variance of the model's estimated target variable is explained by the interaction. It is measured through the $H$ statistic, as in [31], which describes the interaction between a feature $j$ and any others. In particular:

$$H_s^2 = \frac{\sum_{i=1}^{n} \left[\hat{f}(x^{(i)}) - PD_j(x_j^{(i)}) - PD_{-j}(x_{-j}^{(i)})\right]^2}{\sum_{i=1}^{n} \hat{f}^2(x^{(i)})}. \tag{2}$$

$H_s^2$ is scaled in the interval $[0, 1]$, where the absence of interaction corresponds to the null value.

Fig. 10 displays the overall interaction strength of different features. Looking at the order of importance, we observe that USDEUR and NASDAQ have the highest interaction strength, while VIX has the lowest value. The USDEUR's overall interaction explains about 40% of the variance of the estimated CVI.

*SHAP procedure.* To enrich our analysis and compensate for possible "black boxes" originated by the use of advanced ML techniques, we perform the SHAP (SHapley Additive exPlanations) procedure proposed in [37]. In our framework, the process assigns each independent variable an importance value for the CVI prediction. Generally, the method works as follows: consider a finite number $N$ of simplified inputs $x'$ such that $x = h_x(x')$ with $x$ denoting the original inputs. The aim is to find a model that uses the simplified inputs to match the output of the original model $f(x)$. The problem can be described as: $f(x) = \phi_0 + \sum_{j=1}^{N} \phi_j x'_j$, where $\phi_j$ represents the contribution of each feature to the final prediction and $\phi_0$ is a constant value when no inputs are present. In [37], it was proved that the unique solution for all the additive feature attribution methods (i.e. methods which satisfy locally accuracy, missingness and consistency) is:

$$\phi_j = \sum_{z' \subseteq x'} \frac{|z'|!(N - |z'| - 1)!}{N!} [f(z') - f(z' \setminus j)],$$

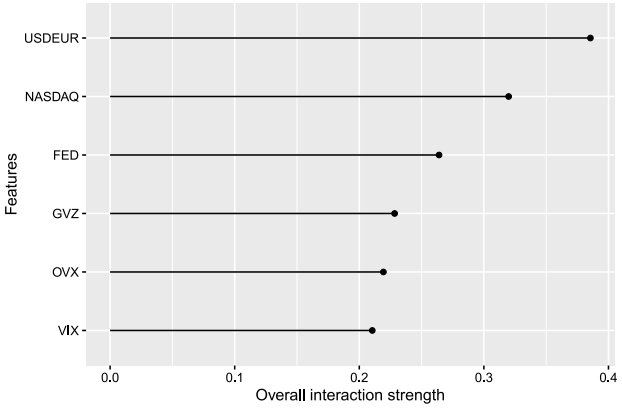



**Fig. 10.** Interactions plots. RF model.

where $z'$ is a subset of $x'$, $|z'|$ is the number of non-zero entries in $z'$ and $z' \setminus j$ means setting $z'_j = 0$, i.e. excluding the $j-$th feature. The values $\phi_j$ are called Shapley values since their expression comes from the classical result of the cooperative game theory due to [38]. The SHAP values represent a unified measure of feature importance and it is defined as the previously described Shapley values $\phi_j$ by placing $f(z') = f(h_x(z')) = \mathbb{E}[f(z)|z_S]$, where $S$ is the set of non-zero entries in $z'$. The SHAP values explain the impact of each feature on the expected model prediction, quantifying how to get from $\mathbb{E}[f(z)]$, without knowing any feature, to the output $f(x)$ (see, e.g., [39]). Fig. 11 illustrates the SHAP contributions of each selected feature to the predictions made by the RF algorithm for specific CVI values. As an example, we present the SHAP contributions for three CVI values: the minimum, the median, and the maximum. Considering that the average contribution is significant across all cases examined, the left panel of Fig. 11(a) illustrates the effects of various features on the CVI minimum value. The most influential factor is FED, which decreases the prediction by 21.574, followed by GVZ and OVX. In contrast, NASDAQ and VIX have the least impact. The second panel, Fig. 11(b), presents the contribution of features to the CVI median value. Unlike the previous case, NASDAQ emerges as the most significant variable, while FED ranks third after USDEUR. Finally, the last panel, Fig. 11(c), examines the CVI maximum value, where GVZ is the most important factor, while the contribution of USDEUR becomes minimal.

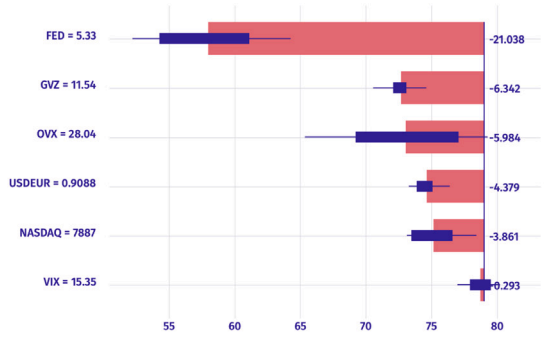
These results suggest that monetary policy (FED) plays a crucial role in low-volatility environments, stock market indices (NASDAQ) become dominant in normal conditions, and volatility-related indicators (GVZ) take over during extreme market turbulence. This dynamic relationship can help in designing adaptive risk management strategies based on prevailing market conditions.

Breiman [32] empirically demonstrates that Random Forest (RF) is relatively robust to noise and outliers. However, assessing the robustness of a RF model can involve various approaches, as robustness encompasses multiple facets of model performance, including stability, generalization, and resistance to overfitting. Regarding the latter, our results indicate that RF does not suffer from overfitting, in contrast to GBM (see Table 4). To assess stability and the model's ability to capture extreme values, Fig. 6 have shown that RF provides a closer fit to the observed density, particularly in higher-value regions. This suggests its effectiveness in capturing extreme data points. Some others techniques to ensure trustworthiness of ML techniques can be found in [40,41]. They introduce the SAFE framework, a set of integrated statistical methods designed to assess Sustainability (or robustness to anomalous data), predictive Accuracy, Fairness (low prediction bias), and Explainability (human interpretation), and applied this procedure to a set of cryptocurrency (Bitcoin) time series. A further observation, in line with [42], should be made regarding the different explainability methods we considered. All methods took only a few minutes to compute on our dataset using a standard laptop, meaning that computational complexity is not a relevant factor for comparison in this case. However, it is important to note that the feature importance measure, typically based on the weighted impurity Gini index, proves to be the most effective in explaining the significance of variables within the RF algorithm. Moreover, as verified in the previous section, RF demonstrates a high level of accuracy. Bearing this in mind, we conclude that, overall, all the proposed procedures yield consistent results under normal conditions, identifying FED, NASDAQ, and USDEUR as the three most influential features.

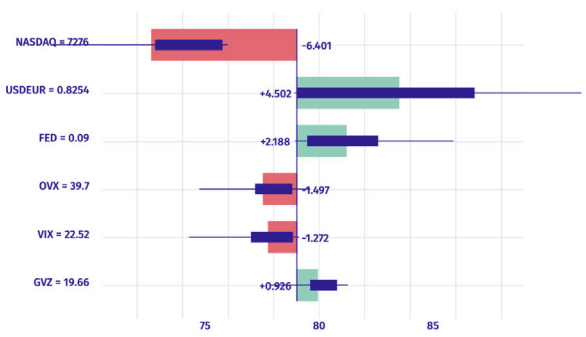
## 5. Conclusions

This paper has explored the potential of Machine Learning (ML) techniques, specifically the Random Forest (RF) algorithm and Gradient Boosting Machine (GBM), to explain the relation between the Crypto Volatility Index (CVI) and six different variables: the Gold Volatility Index (GVZ), Crude Oil Volatility Index (OVX), S&P500 Volatility Index (VIX), the USD/EUR exchange rate (USDEUR), the Federal Reserve interest rate (FED), and the NASDAQ index. Understanding the dynamic accurately is critical for investors who need to anticipate cryptocurrency market volatility trends and assess their relationships with other economic variables to optimize hedging strategies.

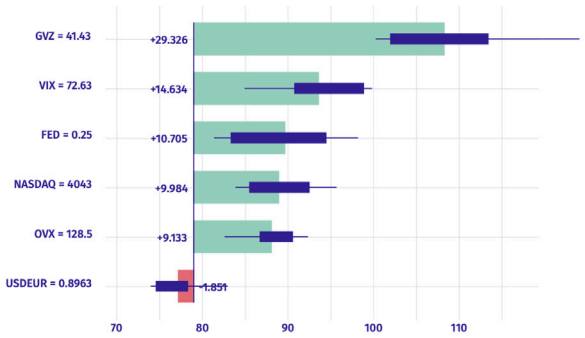
We analyzed daily data from January 1st, 2019, to December 31st, 2023, and justified using ML techniques by demonstrating the potential non-linear relationships among the variables, evaluated through Pearson correlation, distance correlation, and mutual




(a) Minimum CVI value




(b) Median CVI value




(c) Maximum CVI value

**Fig. 11.** Shapley values. RF model.

information. The numerical results have shown a coherent prediction compared to the actual observed values of CVI, with the RF algorithm yielding more accurate results and avoiding overfitting. However, both models have performed well, highlighting the opportunities for further exploration and reliable findings using ML techniques.

Interestingly, the impact of financial volatility variables on CVI prediction has appeared less significant than that of variables representing financial returns. The analysis of non-linear relationships has revealed the limitations of standard linear techniques. While the linear interaction suggests that the NASDAQ index has little impact, it emerges as one of the top three most influential variables when accounting for non-linearity. Our findings have demonstrated how ML techniques can improve cryptocurrency volatility predictions and capture non-linear relationships and evolving variables that linear models may not be able to represent. Although the variables considered are time series, the implemented Random Forest (RF) algorithm lacks temporal awareness and treats each row of the dataset independently. This presents a significant limitation for forecasting, as failing to account for time dependence, the non-stationarity of the series, and variable correlations can result in distorted analyses and misleading conclusions. However, when the objective is explanatory rather than predictive, the analysis can be considered valuable, provided it achieves a good level of accuracy, despite disregarding potential causal validity. This consideration opens the door for future research aimed at overcoming the limit of the proposed model by addressing challenges related to time series analysis. One potential approach to enable the algorithm to effectively handle time series data is the incorporation of time-related and/or lagged variables among the explanatory variables. We plan to develop such a model in a future research work.

The idea behind this article stems from a question: does the cryptocurrency market still maintain its independence from traditional financial markets as initially envisioned?

Cryptocurrencies form part of a complex financial puzzle, exhibiting strong correlations with data from both the real economy and traditional financial systems. This highlighted the need to study them in the context of complex systems. Our proposal to employ machine learning techniques paves the way for further investigations into the construction of complex dynamic models using modern technologies that allow working on big data. Consistent with existing literature, our findings reveal that cryptocurrencies cannot be considered a safe haven for financial investments.

Instead, we conclude that the cryptocurrency market is now embedded within the global dynamics of traditional financial markets. This development suggests that cryptocurrencies can no longer be viewed purely as a hedging instrument for financial investments.

While speculative use of cryptocurrencies is likely to persist, these findings raise pressing questions about financial ethics, which regulatory authorities must address urgently. On the other hand, because of the current increasing prominent role of ML techniques, it could result interesting to further investigate robustness and trustworthiness of ML algorithms as RF and GBM in the context of

cryptofinance. A comprehensive robustness assessment of RF applied to our dataset as well as a deeper investigation of accuracy and fairness as suggested in [40,41], offer research ideas for future works.

## CRediT authorship contribution statement

**Susanna Levantesi:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation. **Gabriella Piscopo:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Formal analysis, Conceptualization. **Alba Roviello:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Investigation, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix. Recursive Feature Elimination (RFE) algorithm

Recursive Feature Elimination (RFE), developed by [43], is one of the most widely used methods for feature selection. Its popularity stems from its simplicity and effectiveness in identifying the most relevant features for predicting a target variable within a training dataset. RFE follows an iterative approach, progressively selecting features by evaluating smaller subsets at each step [44]. The features are ranked according to the sequence in which they are eliminated from the feature space. In the following, we provide the algorithm's steps.

**Algorithm 1** Recursive Feature Elimination (RFE)

**Require:** Feature matrix $X$, Target variable $y$, Subset sizes $f$, Base model $model$, Desired number of features $k$, Cross-validation method $cv$

**Ensure:** Optimal subset of features $F$ with size $k$

$F \leftarrow \{f_1, f_2, \ldots, f_n\}$ ▷ Initialize with all features
**while** $|F| > k$ **do**
    Train the base model $M$ using the current subset of features $F$
    Evaluate model $M$ performance using cross-validation $cv$
    Compute feature importance scores for all $f \in F$
    Remove the least important feature from $F$
    Update the feature set $F$
**end while**
**return** $F$

## Data availability

Data will be made available on request.

## References

[1] A. Brini, J. Lenz, Assessing the resiliency of investors against cryptocurrency market crashes through the leverage effect, Econom. Lett. 220 (2022) 110885.
[2] D.G. Baur, T. Dimpfl, The volatility of Bitcoin and its role as a medium of exchange and a store of value, Empir. Econ. 61 (5) (2021) 2663–2683.
[3] D.G. Baur, J.R. Karlsen, L.A. Smales, A. Trench, Digging deeper-is bitcoin digital gold? A mining perspective, J. Commod. Mark. 34 (2024) 100406.
[4] A. Brini, J. Lenz, A comparison of cryptocurrency volatility-benchmarking new and mature asset classes, Financ. Innov. 10 (1) (2024) 122.
[5] E. Bouri, N. Jalkh, P. Molnár, D. Roubaud, Bitcoin for energy commodities before and after the December 2013 crash: diversifier, hedge or safe haven? Appl. Econ. 49 (50) (2017) 5063–5073.
[6] G. Gajardo, W.D. Kristjanpoller, M. Minutolo, Does bitcoin exhibit the same asymmetric multifractal cross-correlations with crude oil, gold and DJIA as the Euro, great British Pound and Yen? Chaos Solitons Fractals 109 (2018) 195–205.
[7] A.K. Tiwari, I.D. Raheem, S.H. Kang, Time-varying dynamic conditional correlation between stock and cryptocurrency markets using the copula-ADCC-EGARCH model, Phys. A 535 (2019) 122295.
[8] A.H. Dyhrberg, Bitcoin, gold and the dollar–A GARCH volatility analysis, Financ. Res. Lett. 16 (2016) 85–92.
[9] S.H. Kang, R.P. McIver, J.A. Hernandez, Co-movements between bitcoin and gold: A wavelet coherence analysis, Phys. A Stat. Mech. Appl. 120888 (2019).
[10] A. Ghorbel, A. Jeribi, Investigating the relationship between volatilities of cryptocurrencies and other financial assets, Decis. Econ. Finance 44 (2) (2021) 817–843.
[11] N.T. Hung, Asymmetric connectedness among S&P 500, crude oil, gold and Bitcoin, Manag. Financ. 48 (4) (2022) 587–610.
[12] M. Ojaghlou, Co-movements between bitcoin and gold: Multivariate BEKK-GARCH models, in: Computing Intelligence in Capital Market, Springer, 2024, pp. 37–50.
[13] P. Foroutan, S. Lahmiri, Connectedness of cryptocurrency markets to crude oil and gold: an analysis of the effect of COVID-19 pandemic, Financ. Innov. 10 (1) (2024) 68.

[14] A. Attarzadeh, M. Isayev, F. Irani, Dynamic interconnectedness and portfolio implications among cryptocurrency, gold, energy, and stock markets: A TVP-VAR approach, Sustain. Futur. 100375 (2024).
[15] L. Fernandes, J.R.A. Figueirôa, C.M.F. Martins, A.M.F. Martins, The Battle of Informational Efficiency: Cryptocurrencies vs. Classical Assets, Authorea, 2024, http://dx.doi.org/10.22541/au.172797214.45360914/v1.
[16] L. Fernandes, L. Bejan, J.W.L. Silva, E. Bouri, F.H.A. Araujo, The resilience of cryptocurrency market efficiency to COVID-19 shock, 2022, Available At SSRN: https://ssrn.com/abstract=4069490 Or http://dx.doi.org/10.2139/ssrn.4069490.
[17] M.U. Rehman, N. Apergis, Determining the predictive power between cryptocurrencies and real time commodity futures: Evidence from quantile causality tests, Resour. Policy 61 (2019) 603–616.
[18] E. Bouri, R. Gupta, A. Lahiani, M. Shahbaz, Testing for asymmetric nonlinear short-and long-run relationships between bitcoin, aggregate commodity and gold prices, Resour. Policy 57 (2018) 224–235.
[19] V. D'Amato, S. Levantesi, G. Piscopo, Deep learning in predicting cryptocurrency volatility, Phys. A 596 (2022) 127158.
[20] Crypto Briefing, What is the crypto volatility index (CVI)? 2021, Available Online At https://cryptobriefing.com/what-is-the-cvi-how-was-it-developed-what-are-future-plans-for-the-product/.
[21] D.F. Ahelegbey, P. Giudici, NetVIX—A network volatility index of financial markets, Phys. A 594 (2022) 127017.
[22] L. Alessandretti, A. ElBahrawy, L.M. Aiello, A. Baronchelli, Anticipating cryptocurrency prices using machine learning, Complexity 2018 (1) (2018) 8983590.
[23] A. Garcia-Medina, E. Aguayo-Moreno, LSTM-GARCH hybrid model for prediction of volatility in cryptocurrency portfolio, Comput. Econ. 63 (2024) 1511–1542.
[24] E. Akyildirim, A. Goncu, A. Sensoy, Prediction of cryptocurrency returns using machine learning, Ann. Oper. Res. 297 (2021) 3–36.
[25] S. McNally, J. Roche, S. Caton, Predicting the price of bitcoin using machine learning, in: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP, IEEE, 2018, pp. 339–343.
[26] R. Chowdhury, M.A. Rahman, M.S. Rahman, M.R.C. Mahdy, An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning, Phys. A 551 (2020) 124569.
[27] P. Jaquart, S. Köpke, C. Weinhardt, Machine learning for cryptocurrency market prediction and trading, J. Financ. Data Sci. 8 (2022) 331–352.
[28] Y. Liu, Z. Li, R. Nekhili, J. Sultan, Forecasting cryptocurrency returns with machine learning, Res. Int. Bus. Financ. 64 (2023) 101905.
[29] F. Fang, W. Chung, C. Ventre, M. Basios, L. Kanthan, L. Li, F. Wu, Ascertaining price formation in cryptocurrency markets with machine learning, Eur. J. Financ. 30 (1) (2024) 78–100.
[30] L. Breiman, J. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Chapman and Hall/CRC, 1984.
[31] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[32] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[33] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140.
[34] A. Liaw, M. Wiener, Package `Randomforest`, 2024, Available Online At https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.
[35] T.M. Oshiro, P.S. Perez, J.A. Baranauskas, How many trees in a random forest? in: Petra Perner (Ed.), Machine Learning and Data Mining in Pattern Recognition, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN: 978-3-642-31537-4, 2012, pp. 154–168.
[36] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, J. Bus. Econom. Statist. 13 (3) (1995) 253–263.
[37] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[38] L.S. Shapley, A value for n-person games, Contrib. Theory Games 2.28 (1953) 307–317.
[39] G. Babaei, P. Giudici, E. Raffinetti, Explainable artificial intelligence for crypto asset allocation, Financ. Res. Lett. 47 (2022) 102941.
[40] P. Giudici, E. Raffinetti, SAFE artificial intelligence in finance, Financ. Res. Lett. 56 (2023) 104088.
[41] G. Babaei, P. Giudici, E. Raffinetti, A rank graduation box for SAFE AI, Expert Syst. Appl. 259 (2025) 125239.
[42] M.C. Calzarossa, P. Giudici, R. Zieni, An assessment framework for explainable AI with applications to cybersecurity, Artif. Intell. Rev. 58 (5) (2025) 150.
[43] I. Guyon, J. Weston, S. Barnhill, Vapnik V., Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2002) 389–422, http://dx.doi.org/10.1023/A:1012487302797, URL https://doi.org/10.1023/A:1012487302797.
[44] M. Kuhn, K. Johnson, Data pre-processing, in: Applied Predictive Modeling, Springer Science+Business Media, New York, NY, USA, 2013, pp. 27–59, http://dx.doi.org/10.1007/978-1-4614-6849-3_3, URL https://doi.org/10.1007/978-1-4614-6849-3_3.